Chapter XIII

On the Problem of Mining Phrase Definition from Item Descriptions

Hung V. Nguyen, Arizona State University, USA

H. Davulcu, Arizona State University, USA

Abstract

Most search engines do their text query and retrieval using keywords. However, vendors cannot anticipate all possible ways in which shoppers search for their products. In fact, many times, there may be no direct keyword match between a search phrase and descriptions of products that are perfect “hits” for the search. A highly automated solution to the problem of bridging the semantic gap between product descriptions and search phrases used by Web shoppers is developed. By using scalable information extraction techniques from Web sources and a frequent itemset mining algorithm, our system can learn how meanings can be ascribed to popular search phrases with dynamic connotations. By annotating the product databases based on the meanings of search phrases mined by our system, catalog owners can boost the findability of their products.
Introduction

The World Wide Web has made a dramatic transition from its early beginnings as a distributed repository of information that you can browse into a dominant medium for conducting e-commerce. In particular, it has become a mainstream advertising medium for retail goods with online advertising reaching $16 billion in revenue in 2006. Because of this immense commercial power of the Web, the number of retailers, both large and small, who are setting up an online presence and subscribe to search advertising to generate leads to sell their goods through their Web sites, continues to proliferate.

In search advertising, vendors subscribe to triplets of the form <searchphrase, product-url, bid>. For example, a shoe store may subscribe to advertise its “NIKE Airmax 180” product as “running shoes” by specifying the triplet <runningshoes, NIKEAirmax180, $.50>. This triplet indicates that whenever a Web site contains “running shoes” related information, this vendor would like to list its “NIKE Airmax 180” product and agrees to pay 50 cents per click. Nevertheless, vendors cannot anticipate all possible ways in which to advertise their products. Advertising such as Google’s AdSense utilize sense-disambiguation and keyword matching to place ads on highly relevant Web pages. However, there may be a semantic gap between the search phrases used by vendors and the way the relevant content is presented in the Web pages. For example, if a Web page mentions “stable light shoes,” a smart advertising algorithm should be able to detect the relationship between “stable light shoes” and “running shoes,” thus, relieving the advertiser from the burden of specifying all potential relevant search-phrases for “running shoes.”

Keyword searching is the most common form of product search on the Web. Most search engines do their text query and retrieval using keywords. The average keyword query length is under three words (2.2 words) (Cutting & Douglas, 1997). Recent research (Andrews, 2003) found that 40% of e-commerce companies rate their search tools as “not very useful” or “only somewhat useful.” Further, a review of 89 sites (Andrews, 2003) found that 75% have keyword search engines that fail to retrieve important information and put results in order of relevance; 92% fail to provide guided search interfaces to help offset keyword deficiencies (Andrews, 2003), and 7 out of 10 Web shoppers were unable to find products using the search engine, even when the items were stocked and available.

The defining problem: Vendors cannot anticipate all possible ways in which shoppers search for their products. In fact, many times, there may be no direct keyword match between a search phrase and descriptions of products that are perfect “hits” for the search. For example, if the shopper uses “motorcycle jacket,” then, unless the publisher knows that every “leather jacket” is a “motorcycle jacket,” it cannot produce all matches for the user’s search. Thus, for certain phrases, there is a semantic gap between the search phrase used and the way the corresponding match-
A Fuzzy Clustering Model for Fuzzy Data with Outliers
www.igi-global.com/chapter/fuzzy-clustering-model-fuzzy-data/67480?camid=4v1a